VBEST NOTES

A LEVEL CIE

A2 STATISTICS 2 (9709)

www.vbest.edu.my

Statistics

- [ ] Poisson distribution
- [ ] Linear combinations of random variables
- [ ] Continuous random variables
- [ ] Sampling and estimation
- [ ] Hypothesis testing

Chapter 1 : Poisson Distribution

a) Poisson distribution

$X \sim Po(\lambda)$, where λ represents constant average rate

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, 3, \ldots \text{ to infinity}$$

Eg : $X \sim Po(2)$

i) $P(X = 4) : \frac{(e^{-2})(2^4)}{4!} = 0.9022$

ii) $P(X > 4) : 1 - e^{-2}\left(1 + 2 + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!}\right) = 0.0527$

mean and variance are both the λ value

b) Poisson distribution as an approximation to the binomial distribution

$X \sim B(n,p)$; $X \sim Po(\lambda)$
$\mu = np$ $\quad\quad \mu = \lambda$
$\therefore np = \lambda$

Eg : $X \sim B(80, 0.10)$

$80 \times 0.10 = 8$ $\quad X \sim Po(8)$

$$P(X < 2) = e^{-8}\left(1 + 8 + \frac{8^2}{2!}\right) = 0.01375$$

X can be approximated by $Po(np)$ if
n is large, p is small and $np > 5$

c) Normal distribution as an approximation to Poisson distribution

$X \sim Po(\lambda)$; $X \sim N(\mu, \sigma^2)$
Var = npq $\quad\quad$ Var = σ^2
Can be approximated if $\lambda > 15$ (large)

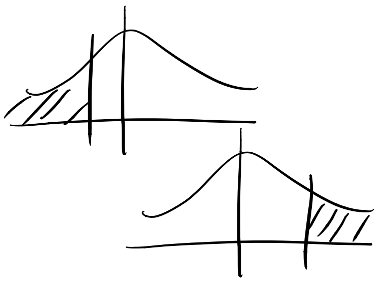
Continuity correction is required since you are using a continuous distribution as an approximation to a discrete distribution

Eg : $X \sim Po(30)$
$P(X \leq 20)$

$$P\left(Z \leq \frac{20.5 - 30}{\sqrt{30}}\right)$$

$P(Z \leq -1.734)$

$p = 1 - 0.9586 = 0.0414$



Chapter 2 : Linear combinations of random variables

a) Mean and variance

$E(aX) = a\,E(X)$ $\quad$ $E(aX + b) = a\,E(X) + b$

$Var\,(aX) = a2\,Var(X)$ $\quad$ $Var\,(aX + b) = a2\,Var(X)$ variance is not affected by +b

Eg : The random variable X has mean 20 and variance 4.

i) $6X + 1$

Mean = 121 and Variance = 25

ii) $3X - 2$

Mean = 58 and Variance = 10

b) Sums and differences of independent random variables

$E(aX \pm bY \pm c) = a\,E(X) \pm b\,E(Y) \pm c$

$Var\,(aX \pm bY \pm c) = a^2\,Var(X) + (-b)^2\,Var(Y)$

$= a^2\,Var(X) + b^2\,Var(Y)$

Variance is not affected by c and is always added

Eg : X and Y are independent variables such that E(X) = 8, Var (X) = 2, E(Y) = 10 and Var(Y) = 3. Find the mean and variance of $3X + 2Y$

Mean : $3(8) + 2(10) = 44$

Variance : $3^2\,(2) + 2^2(3) = 30$

Independant observation

If random variables are independent there is no square for variance

$E(X1 + X2) = E(X1) + E(X2)$
$= E(X) + E(X)$
$= 2\,E(X)$

$Var(X1 + X2) = Var(X1) + Var(X2)$
$= Var(X) + Var(X)$
$= 2\,Var(X)$

	$X_1 + X_2$	$2X$
Mean	$E(X_1 + X_2) = 2\,E(X)$	$E(2X) = 2\,E(X)$
Variance	$Var(X_1 + X_2) = 2\,Var(X)$	$Var(2X) = 2^2\,Var(X)$

c) The sum of independent Poisson variables

If $X \sim Po(\lambda_2)$ and $Y \sim Po(\lambda_2)$ then $X + Y \sim Po(\lambda_1 + \lambda_2)$

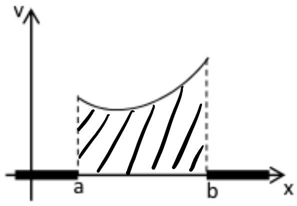
Mean and variance are both $\lambda_1 + \lambda_2$

Chapter 3 : Continuous Random Variable

a) Probability density function

$$P(a < X < b) = \int_a^b f(x)dx$$

Area under the graph = 1



b) Mean and variance

Mean : $$\int_{-\infty}^{\infty} x\, f(x)\, dx$$

Variance : $$\int_{-\infty}^{\infty} x^2\, f(x)\, dx - \mu^2$$

c) Median and quartiles

If the continuous random variable X is defined by its probability density function for $a \le x \le b$ then

$\int_a^m f(x) = \frac{1}{2}$ for median

$\int_a^{q1} f(x) = \frac{1}{4}$ for lower quartile

$\int_a^{q3} f(x) = \frac{3}{4}$ for upper quartile

Chapter 4 : Sampling and estimation

a) Sample mean and variance $\overline{X}$

$$E(\bar{X}) = \mu$$

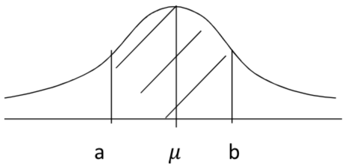
$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

The standard deviation of the distribution of sample is known as standard error $\frac{\sigma}{\sqrt{n}}$

b) Central limit theorem

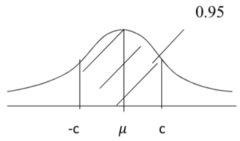
- If X is normally distributed then $\overline{X}$ will also be normally distributed.
- If X is not normally distributed but n is large then the distribution of $\overline{X}$ can be approximately normally distributed.
- When the normal distribution is used as an approximation to a discrete distribution a continuity correction is needed.

c) Confidence interval



A and B are known as confidence limits

95% confidence interval for μ if $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

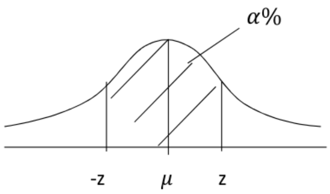


If 100 samples are taken, 95 of the samples are said to **contain** the population mean or

P (interval contains μ) = 0.95 .

The α% of confidence interval is $\left(\bar{x} - z\frac{\sigma}{\sqrt{n}},\ \ \bar{x} + z\frac{\sigma}{\sqrt{n}}\right)$

The width of the α% confidence interval is $2 \times z\frac{\sigma}{\sqrt{n}}$



Chapter 5 : Hypothesis testing

Null hypothesis H_0 : hypothesis that we assume to be correct unless proven otherwise

Alt hypothesis H_1 : tells us the value of population parameter if our assumption is shown to be wrong

Critical region : the range of values of the test centre state that need you to rejecting H_0

Critical value : boundaries of the critical region

Level of significance : threshold probability that varies depending on nature of the problem

a) Critical region

Range of values for which you reject the null hypothesis is known as critical region

Eg :

For a 5% significance level

i) *Upper tail:* critical value at *c*,

the critical region consists of values greater than or equal to *c* such that $P(X \geq c) < 0.05$

ii) *Lower tail:* critical value at *c*,

the critical region consists of values less than or equal to *c* such that $P(X \leq c) < 0.05$

ii) *Two-tail:* critical value at c_1 and c_2,

the critical region consists of values greater than or equal to *c* such that $P(X \geq c_1) < 0.025$

and values less than or equal to *c* such that $P(X \leq c_2) < 0.025$

b) Hypothesis testing

Step 1: Define the variable and its distribution

Step 2: State the null hypothesis, H0 and the alternative hypothesis, H1

Step 3: State the rejection rule (either a probability statement or critical region)

Step 4: Find whether the test value lies in the critical region by calculating the probability and comparing to significant level.

Step 5: Make your conclusion in statistical terms

One tailed hypothesis

$H_o: \theta = m$

$H_1: \theta > m$

Reject H_o if $P(X \geq x) \leq \propto \%$

$H_o: \theta = m$

$H_1: \theta < m$

Reject H_o if $P(X \leq x) \leq \propto \%$

Two-tailed hypothesis

$H_o: \theta = m$

$H_1: \theta \neq m$

Reject if $P(X \geq x) \leq \frac{1}{2} \propto$ or $P(X \leq x) \leq \frac{1}{2} \propto$

c) Type I and type II errors

Type I error is made when true H_0 hypothesis is rejected

Type II error is made when false H_1 is rejected

*To find the probability of a Type II error you must be given a specific value for alternative H_1

1st Edition : Zhi Yee

Producer : Mr. Sai Mun